

Exercise: Generating Test Data for a Database with ChatGPT

Objective: Create realistic test data for the **Accounts** table schema.

Key Steps:

1. **Provide Detailed Schema:** Give ChatGPT the **Accounts** table schema.
2. **Craft Precise Prompts:** Include requirements for realistic and hierarchically consistent data.
3. **Generate and Review Data:** Use ChatGPT to generate data, then format it into SQL **INSERT** statements.

Outcome: Learn to use ChatGPT for efficient test data generation, enhancing database development workflows.

Duration: 5 to 10 minutes

Background

In the realm of database development and testing, generating realistic test data is a critical yet often challenging task. Accurate and comprehensive test data ensures that database systems are robust, efficient, and capable of handling real-world scenarios. Manual data generation is not only time-consuming but also prone to errors, making it difficult to cover all possible test cases.

Automated test data generation, especially with the help of advanced tools like ChatGPT, offers a solution to these challenges. By leveraging AI, developers can quickly generate large volumes of realistic data that adhere to specific schema requirements. This not only speeds up the testing process but also improves the quality and reliability of database applications.

Objective

Generate test data for the **Accounts** table based on the following schema, which can be copied from any RDBMS like Oracle, SQL Server, Informix, etc.:

```
CREATE TABLE [dbo].[Accounts](
    [Account_ID] [bigint] NOT NULL,
    [Account_Name] [varchar](255) NULL,
    [Modified_Time] [datetimeoffset](7) NULL,
    [Description] [varchar](max) NULL,
    [Billing_Street] [varchar](max) NULL,
    [Created_Time] [datetimeoffset](7) NULL,
    [Website] [varchar](max) NULL,
    [Employees] [int] NULL,
    [Parent_Account] [varchar](max) NULL,
    [Billing_City] [varchar](255) NULL,
    [Phone] [varchar](255) NULL,
    [Billing_Country] [varchar](255) NULL,
    [Billing_State] [varchar](255) NULL)
```

Prompt

Based on the following SQL Server table schema for 'Accounts', please generate 10 rows of test data. Give me the Insert statements so I can insert the information into a table.

```
CREATE TABLE [dbo].[Accounts](
    [Account_ID] [bigint] NOT NULL,
    [Account_Name] [varchar](255) NULL,
    [Modified_Time] [datetimeoffset](7) NULL,
    [Description] [varchar](max) NULL,
    [Billing_Street] [varchar](max) NULL,
    [Created_Time] [datetimeoffset](7) NULL,
    [Website] [varchar](max) NULL,
    [Employees] [int] NULL,
    [Parent_Account] [varchar](max) NULL,
    [Billing_City] [varchar](255) NULL,
    [Phone] [varchar](255) NULL,
    [Billing_Country] [varchar](255) NULL,
    [Billing_State] [varchar](255) NULL)
```

Review the information provided by the model, are the results correct?

In most cases, providing a poor prompt will set you up to receive poor results.

Playground

Manual evaluationDeploy to a web appImportExportView CodePrompt samplesCustomize in prompt flow

System messageAdd your dataExamples

Ask questions about your own data. Your data is stored securely in your Azure subscription. [Learn more about how your data is protected.](#)

Select project index *
Select project index
+ Add your data

Replay chatClear chatPlayground SettingsShow JSON

[Employees] [int] NULL,
[Parent_Account] [varchar](max) NULL,
[Billing_City] [varchar](255) NULL,
[Phone] [varchar](255) NULL,
[Billing_Country] [varchar](255) NULL,
[Billing_State] [varchar](255) NULL

Here are the INSERT statements for generating 10 rows of test data for the 'Accounts' table:

INSERT INTO [dbo].[Accounts] ([Account_ID], [Account_Name], [Modified_Time], [Description], [Billing_Street],
[Created_Time], [Website], [Employees], [Parent_Account], [Billing_City], [Phone], [Billing_Country], [Billing_State])
VALUES (1, 'Account 1', '2022-01-01 12:00:00', 'Description 1', 'Street 1', '2022-01-01 12:00:00', 'www.account1.com', 10,
'Parent Account 1', 'City 1', '123-456-7890', 'Country 1', 'State 1');

INSERT INTO [dbo].[Accounts] ([Account_ID], [Account_Name], [Modified_Time], [Description], [Billing_Street],
[Created_Time], [Website], [Employees], [Parent_Account], [Billing_City], [Phone], [Billing_Country], [Billing_State])
VALUES (2, 'Account 2', '2022-01-02 12:00:00', 'Description 2', 'Street 2', '2022-01-02 12:00:00', 'www.account2.com', 20,
'Parent Account 2', 'City 2', '234-567-8901', 'Country 2', 'State 2');

INSERT INTO [dbo].[Accounts] ([Account_ID], [Account_Name], [Modified_Time], [Description], [Billing_Street],
[Created_Time], [Website], [Employees], [Parent_Account], [Billing_City], [Phone], [Billing_Country], [Billing_State])
VALUES (3, 'Account 3', '2022-01-03 12:00:00', 'Description 3', 'Street 3', '2022-01-03 12:00:00', 'www.account3.com', 30,
'Parent Account 3', 'City 3', '345-678-9012', 'Country 3', 'State 3');

Type user query here. (Shift + Enter for new line)

Confidential - Actionable Data Analytics Pty Ltd (ABN 79 649 102 536) - 2 / 5

Modify the prompt to include better directions.

Based on the following SQL Server table schema for 'Accounts', please generate 10 rows of test data, the information needs to look real (with correct names, addresses, cities, etc) but don't use real information.

Make sure the hierarchy between city, state and country exists and it is real.

Give me the Insert statements so I can insert the information into a table.

```
CREATE TABLE [dbo].[Accounts](
    [Account_ID] [bigint] NOT NULL,
    [Account_Name] [varchar](255) NULL,
    [Modified_Time] [datetimeoffset](7) NULL,
    [Description] [varchar](max) NULL,
    [Billing_Street] [varchar](max) NULL,
    [Created_Time] [datetimeoffset](7) NULL,
    [Website] [varchar](max) NULL,
    [Employees] [int] NULL,
    [Parent_Account] [varchar](max) NULL,
    [Billing_City] [varchar](255) NULL,
    [Phone] [varchar](255) NULL,
    [Billing_Country] [varchar](255) NULL,
    [Billing_State] [varchar](255) NULL)
```

Review the results, is the test data more accurate?

...

Sure! Here are the INSERT statements for generating 10 rows of test data for the 'Accounts' table with realistic but fictional information:

```
INSERT INTO [dbo].[Accounts] ([Account_ID], [Account_Name], [Modified_Time], [Description], [Billing_Street],
[Created_Time], [Website], [Employees], [Parent_Account], [Billing_City], [Phone], [Billing_Country], [Billing_State])
VALUES (1, 'ABC Corporation', '2022-01-01 12:00:00', 'A fictional corporation specializing in software development',
'123 Main Street', '2022-01-01 12:00:00', 'www.abccorp.com', 100, NULL, 'New York City', '555-123-4567', 'United
States', 'New York');
```

```
INSERT INTO [dbo].[Accounts] ([Account_ID], [Account_Name], [Modified_Time], [Description], [Billing_Street],
[Created_Time], [Website], [Employees], [Parent_Account], [Billing_City], [Phone], [Billing_Country], [Billing_State])
VALUES (2, 'XYZ Industries', '2022-01-02 12:00:00', 'A leading manufacturer of industrial equipment', '456 Oak Avenue',
'2022-01-02 12:00:00', 'www.xyzindustries.com', 500, NULL, 'Los Angeles', '555-234-5678', 'United States', 'California');
```

```
INSERT INTO [dbo].[Accounts] ([Account_ID], [Account_Name], [Modified_Time], [Description], [Billing_Street],
[Created_Time], [Website], [Employees], [Parent_Account], [Billing_City], [Phone], [Billing_Country], [Billing_State])
VALUES (3, 'Acme Corporation', '2022-01-03 12:00:00', 'A global conglomerate with interests in various industries', '789
Maple Street', '2022-01-03 12:00:00', 'www.acmecorp.com', 1000, NULL, 'Chicago', '555-345-6789', 'United States',
'Illinois');
```

```
INSERT INTO [dbo].[Accounts] ([Account_ID], [Account_Name], [Modified_Time], [Description], [Billing_Street],
[Created_Time], [Website], [Employees], [Parent_Account], [Billing_City], [Phone], [Billing_Country], [Billing_State])
```

Relationships

In many cases, you will want to generate test data that is related to the information in a different table. For those scenarios, you can use the "historical" information in the chat to continue asking for more information.

Example generating contacts:

```
generate 3, 5 or 10 test contacts per each account. Taking into account the following Contacts schema. The information needs to look real. Generate the insert statements.
```

```
CREATE TABLE [dbo].[Contacts](
    [First_Name] [varchar](255) NULL,
    [Email] [varchar](255) NULL,
    [Last_Name] [varchar](255) NULL,
    [Phone] [varchar](255) NULL,
    [Account_ID] [bigint] NULL,
    [Modified_Time] [datetimeoffset](7) NULL,
    [Created_Time] [datetimeoffset](7) NULL,
)
```

Sure! Here are the INSERT statements for generating test contacts for each account in the 'Contacts' table:

For Account ID 1:

```
INSERT INTO [dbo].[Contacts] ([First_Name], [Email], [Last_Name], [Phone], [Account_ID], [Modified_Time],
[Created_Time])
VALUES ('John', 'john@example.com', 'Doe', '555-111-1111', 1, '2022-01-01 12:00:00', '2022-01-01 12:00:00');
INSERT INTO [dbo].[Contacts] ([First_Name], [Email], [Last_Name], [Phone], [Account_ID], [Modified_Time],
[Created_Time])
VALUES ('Jane', 'jane@example.com', 'Smith', '555-222-2222', 1, '2022-01-02 12:00:00', '2022-01-02 12:00:00');
INSERT INTO [dbo].[Contacts] ([First_Name], [Email], [Last_Name], [Phone], [Account_ID], [Modified_Time],
[Created_Time])
VALUES ('Mike', 'mike@example.com', 'Johnson', '555-333-3333', 1, '2022-01-03 12:00:00', '2022-01-03 12:00:00');
```

For Account ID 2:

```
INSERT INTO [dbo].[Contacts] ([First_Name], [Email], [Last_Name], [Phone], [Account_ID], [Modified_Time],
[Created_Time])
VALUES ('Sarah', 'sarah@example.com', 'Wilson', '555-444-4444', 2, '2022-01-04 12:00:00', '2022-01-04 12:00:00');
INSERT INTO [dbo].[Contacts] ([First_Name], [Email], [Last_Name], [Phone], [Account_ID], [Modified_Time],
[Created_Time])
VALUES ('Chris', 'chris@example.com', 'Brown', '555-555-5555', 2, '2022-01-05 12:00:00', '2022-01-05 12:00:00');
INSERT INTO [dbo].[Contacts] ([First_Name], [Email], [Last_Name], [Phone], [Account_ID], [Modified_Time],
[Created_Time])
VALUES ('Emily', 'emily@example.com', 'Taylor', '555-666-6666', 2, '2022-01-06 12:00:00', '2022-01-06 12:00:00');
```

For Account ID 3:

```
INSERT INTO [dbo].[Contacts] ([First_Name], [Email], [Last_Name], [Phone], [Account_ID], [Modified_Time],
```

Exercise Completion

In this exercise, you learned to use ChatGPT for generating realistic test data for a database schema. By providing a detailed schema and crafting precise prompts, you discovered how to create accurate and hierarchically consistent SQL **INSERT** statements. This skill is essential for enhancing your database development and testing workflows.